

PREDICTING THE ACTIVITIES OF AN INDIVIDUAL OR GROUP USING MINIMAL INFORMATION

Background Of The Invention

5

Field of the Invention

The present invention relates to a method and system for predicting the future behavior of an individual or group engaged in a particular type of activity when there is little or no information on previous behavior of that specific individual or group under comparable conditions.

10

Description of Related Art

An age-old question faced by advertising and marketing professionals is how to ensure that their advertising and marketing materials reach the largest possible audience that is favorably disposed to purchase the products or services being promoted, so as to maximize the effectiveness of an advertising campaign. Clearly, advertising via mass media, such as in newspapers and magazines having wide circulation or on network television programming, will cause the message to be delivered to the largest number of consumers in the shortest possible period of time. However, it is very likely that only a miniscule percentage of the persons to whom these materials are exposed at any given time would be attentive to the contained message or would be in the market for the product or service in question, at the exact moment of exposure. Therefore, such approaches are both inefficient and costly, since the cost is generally based on the number of persons to whom the advertising will be delivered, irrespective of their potential interest.

15

20

25

30

Clearly, a more targeted approach would be more cost-efficient, i.e., one that is limited only to consumers who are likely to be favorably predisposed to the products or services being promoted. Although the marketing messages

would be viewed by fewer persons, yet, from a statistical perspective, a higher percentage of those persons would be likely to be induced to take positive action in response to that message. This results in a markedly more effective use of an advertising budget. A consumer's perceived predisposition could be based, for example, on his/her: being a member of a particular socio-economic class; having a relevant occupation or hobby; living in a certain geographic area; having a family of a certain size; being a graduate of a particular type of school; having a certain ethnic background; subscribing to certain periodicals; or having bought similar products or services in the recent past.

Therefore, over the years, retailers and their advertising affiliates have spent considerable time and money in acquiring, analyzing and categorizing information from millions of individual consumers. Information is acquired using traditional methods such as in-person and telephone surveys, and, in more recent years, through more sophisticated methods such as monitoring use of customer "loyalty cards" in connection with purchases, and monitoring web-surfing activity over various on-line services that access the Internet using "cookies" or comparable data-gathering mechanisms. Through well-known statistical and probability-based modeling techniques, this archive of pertinent information can be analyzed and processed in various ways to identify an individual consumer's preferences or predispositions to become engaged in certain types of activities. By comparing information collected from individuals sharing common attributes, group preferences or predispositions similarly can be established.

This information then can be used to deliver targeted advertising content to those individual consumers, or similarly oriented groups of consumers, who are most likely to be influenced by the advertising message, whether by means of traditional direct mailings, telemarketing programs, or real-time banner

advertisements visible on a computer monitor screen during a particular user's on-line surfing activities, or television ads received during the normal broadcast or during interactive sessions. This results in a more systematic, more controlled delivery of content over the life span of a particular advertising campaign.

5

However, to date relatively little, if any, targeted content delivery could be made to a given individual unless a meaningful amount of data on that individual's past activity, within a particular environment or setting, had been compiled. This is especially a problem in the case of on-line delivery of advertising messages, e.g., when the user is visiting a specific Internet web site, since there is a relatively limited window of opportunity during which to reach the on-line user before he/she either leaves the web site in question to visit another web site, or logs off entirely from the on-line service that provides the access to the web sites. If a person is a first-time visitor to a particular on-line environment (e.g., a particular web site), and if there is no past history of either how that person is known to act within that environment, or how that person reacts to situations that are typical of those found in that environment, then targeted content delivery, if attempted at all, would need to be based on broad assumptions that may not in fact be applicable to that person. Some reports estimate that as many as 80% of all visitors to a web site are "unknown" to that web site, i.e., the web site has no data whatsoever on those visitors.

10

15

20

25

Therefore, it would be desirable to provide a technique by which to identify tendencies or preferences of a particular user of an information delivery service, based on similarities between the user's present activities in connection with that service and the characteristics exhibited by other users of the same or similar information delivery services.

It would also be desirable to use the identified tendencies or preferences to anticipate or predict the user's future behavior in connection with the information delivery service.

- 5 It would also be desirable to use the identified tendencies or preferences to deliver targeted informational content to the user based on the identified tendencies or preferences.

Brief Summary Of The Invention

One embodiment of the present invention is directed to a technique for predicting the behavior of a current user of an interactive service. Each activity in which the current user participates while engaged with the interactive service is identified, as well as the conditions surrounding each activity. A first collection of data is accessed that reflects (i) cumulative activities in which other users have participated, (ii) conditions surrounding the other users' cumulative activities, and (iii) patterns of behavior exhibited by the other users derived from their participation in such cumulative activities. The current user's identified activities and surrounding conditions are compared with the other users' cumulative activities and surrounding conditions, to identify similarities. Finally, a pattern of future behavior is attributed to the current user, based on the identified similarities and on the other users' patterns of behavior.

In another embodiment of the invention, additionally there is a second collection of data that reflects (i) the types of activities (e.g., categorized by informational content) that are available via the interactive service and (ii) information about each type of activity that distinguishes it from other types of activities. The embodiment attributes a pattern of behavior to the current user by selecting one or more types of activities from the list of all available types of activities, as being the one(s) in which the current user is most likely to participate during his/her engagement with the interactive service.

In further embodiment of the invention, the behavior of a user of an interactive service can be predicted during a particular period of engagement with that interactive service, even in the absence of a pre-existing collection of data reflecting the cumulative activities of multiple other prior users. This embodiment (a) identifies activities, or types of activities, in which the user participates during that period of engagement and the conditions surrounding

each such activity (b) identifies activities, or types of activities, in which other users contemporaneously participate during the same period of engagement and the conditions surrounding each such activity and (c) compiles in real time a collection of data that reflects the cumulative activities of the other users and their surrounding conditions. Then, patterns of behavior are discerned, in real time and on an ongoing basis, from this expanding and evolving collection of data, and data reflecting these patterns of behavior are incorporated into the collection of data. A comparison is made between (i) the user's present activities and surrounding conditions and (ii) the collection of data, to identify similarities. Finally, there is attributed to the user a pattern of future behavior based on such similarities and on the previously determined patterns of behavior.

Based on these predictions of behavior, targeted informational content can then be delivered to the user, or to groups of similarly oriented users, that is relevant to the user's perceived areas of interest.

Embodiments of the invention are based in part on the recognition that people in general behave in very predictable manners. Their activities follow patterns that are easily discernable from one another. Learning how to identify the attributes of a particular pattern allows one to forecast with a high degree of certainty what an individual's next action is likely to be. For example, in the case of visits to Internet web sites, it is generally known that most users participate in only a limited number of activities while on the Web and that they participate in only one activity at a time. Also, a typical user will generally spend a certain amount of time on that one activity before moving on to another activity. Certain groups of people will behave the same with regard to the sequence in which they move from one activity to another and to the amount of time participating in each activity. These are the emerging set of patterns of behaviors

across users that can be used to compare to an individual user, even if he/she is relatively unknown.

5 These and other features of various embodiments of the present invention will become readily apparent from the following detailed description, to be read in conjunction with the appended drawings. The description and drawings are intended to be merely illustrative in nature, and not to restrict or limit the scope of the invention as defined by the claims.

Brief Description Of The Drawings

FIG. 1 is a block diagram illustrating a representative network in which the invention can be implemented;

5 FIG. 2 is a block diagram illustrating the functional components of a behavior-predicting system configured in accordance with an embodiment of the present invention;

10 FIG. 3 is a flow diagram illustrating typical operation of a system configured in accordance with an embodiment of the present invention; and

 FIG. 4 is a block diagram illustrating the functional components of a behavior-predicting system in accordance with an alternative embodiment.

Detailed Description Of Preferred Embodiments

The present invention is directed to predicting the probable or likely behavior of an individual person (or a group of persons who are related according to certain criteria) interacting with an information access service, based primarily on the behavioral traits or preferences demonstrated by other persons (or other groups) who have transacted with that same or a similar service, as opposed to being based exclusively on observations of, and analysis of, historical actions of the specific individual (or specific group) being monitored. In a specific embodiment that operates in an Internet-based environment, the types of Internet web sites that the user in question is most likely to subsequently visit can be determined during a particular session of web-surfing and the amount of time that the user is likely to spend at such web sites.

Embodiments of the invention are particularly useful in conjunction with a targeted content delivery system, which allows the web site to present to the individual user being observed advertisements, promotional messages or other informational content that is relevant to the predicted behavior of that user, and to which that user is most likely to be responsive. An example of such a targeted content delivery system is the one disclosed in U.S. patent application Serial No. 09/558,755 filed April 21, 2000 in the name of the assignee of the present application, entitled "Method and System for Web User Profiling and Selective Content Delivery", and which is incorporated herein by reference. Such a content delivery system can be configured to initiate a content delivery in response to the user's behavioral pattern or to his/her tendencies or traits, as determined in accordance with the present invention.

FIG. 1 shows a representative network in which the inventive prediction system can be implemented. The network includes a plurality of client machines operated by various individual users. The client machines connect to

multiple servers 12 via a communication channel 14, e.g., the Internet. In the case of the Internet, the servers 12 are known as web servers and they are selectively accessible by various clients. The web servers 12 support so-called "web sites" that consist of one or more "pages", i.e., files in various formats such as text, graphics, images, sound, video, etc. A network path to a web site generated by its associated web server is identified by a Uniform Resource Locator (URL).

One example of a client machine 10 is a personal computer, including a central processing unit (CPU), operating system and associated application programs, memory, input devices such as a keyboard and/or a mouse, and output devices, typically including a visual display unit. The screen of the display unit can be used to present a graphical user interface (GUI) by which the user interacts with the client machine, and in turn with the Internet network of servers, in a well-known manner.

Client machines 10 typically employ browsers, which are well-known software applications used to access the web servers 12, such as, for example, Netscape Navigator™ and Microsoft Internet Explorer™. Client machines 10 can access web servers 12 through an Internet service provider (ISP) such as, for example, America Online™. Illustrated in FIG. 1 are multiple ISP "point-of-presence" (POP) systems, each of which includes a POP server 16 linked to a group of client machines 10 for providing access to the Internet via the ISP's local area network. Each POP server 16 can accept URL page requests from one or more client machines 10, forward such requests to the Internet for accessing the requested URLs and distribute targeted informational content back to the users of such client machines. In addition, each POP server can facilitate the creation of a record of all such URL page requests for analysis in accordance with the teachings of the present invention.

The prediction system further includes a remote master server 18 linked to each POP server 16, preferably through the Internet 14. The computer software used to implement this preferred embodiment can be resident on the master server 18 (as in the FIG. 2 example described below), on the various POP servers 16, on the various client machines 10, or on a combination of these entities; such software can be distributed from the master server 18 to, and installed on, the POP servers 16 and the client machines 10 via the Internet 14, in a conventional manner.

FIG. 2 illustrates in greater detail the functional components of a particular embodiment of a prediction system in accordance with the present invention. (For simplicity, only a single client machine 10 and its associated POP server 16 are shown.) A data stream 20 associated with a particular current user whose activities are being monitored and who is a potential recipient of some targeted informational content, is input via the POP server into an Analytical Module (AM) 22 operative on the Master Server 18. This data stream 20 contains (a) the URLs requested by the current user and therefore the associated web sites that such user has elected to visit during his/her current web-surfing session, and (b) the amount of time that the user has spent investigating each web site. This data stream is supplied by a software routine contained within a Content Delivery System (CDS) software client 24 installed in the client machine that detects the key strokes, mouse clicks or other inputs generated by the user during a visit to a particular web site or to communicate the user's desire to visit another web site. (The CDS software client 24 works in conjunction with a Content Delivery System 34 described later in connection with FIG. 2.) Alternatively, this data stream could be supplied by any one of various conventionally known technologies that reside on, or otherwise interface with, each client machine 10 to monitor a user's on-line activity, e.g., "cookies".

A second data stream 26 is provided to the AM 22 from a database 28 that contains files reflecting the historical activities of many other visitors to web sites, whether those web sites have been accessed via the ISP in question or via another ISP. Each such data file contains information that in some way identifies the corresponding user, usually by an anonymous identification number as well as by generic characteristics or attributes associated with that individual, e.g., age, gender, occupation, city or state of residence, amount of income, family size, etc. This second data stream also contains information reflecting the various types of web sites that have been visited by the other users as well as the amount of time spent at each type of web site. Generally, this information would have been compiled from an analysis of these users' click stream data, in a manner such as is described in the previously mentioned patent application Serial No. 09/558,755. Alternatively, some of this information could have been obtained by a voluntarily disclosure by the other users, e.g., by filling out an on-line questionnaire in order for those users to obtain certain privileges or benefits from the ISP or from the operator of a particular web site.

Finally, the second data stream contains information in the form of behavioral patterns, trends or preferences exhibited by prior users that share one or more attributes or groupings of attributes, derived from a conventional statistical analysis of all the compiled raw data on prior user attributes, types of web sites visited and time duration of visits. For example, based on a statistical analysis of the web-surfing patterns of millions of users (whether via this particular ISP or other ISPs), it might be determined that the vast majority of over-50 males living in the New England states and having an annual income exceeding \$100,000, who visit a travel-oriented web site for more than ninety seconds during a particular on-line session will also visit a web site that offers books for sale and will stay at that subsequent web site for an average of 3.5 minutes.

Naturally, the larger the amount of data collected about other users, both as to the number of users and the amount of data per user, the more detailed the behavioral patterns that can be discerned and the more accurate predictions that can be successfully made by the present invention. Therefore, the database 28 and consequently the second data stream 26 are continually updated to reflect ongoing visits to web sites by other users, or even visits by newly added users, contemporaneous with the current user conducting his/her present on-line session.

It should be pointed out that although the database 28 has been characterized as containing data reflecting the activities of other users, it also could contain data reflecting certain previous activities of the user who currently is being observed, and whose behavior is being predicted. For example, even though the current user may be visiting a particular web site (or type of web site) for the first time, and therefore that web site has no specific knowledge of that user's past activities and/or behavioral patterns, nevertheless there could be considerable accumulated data on that user's activities and behavioral patterns derived from an observation of his/her activities at other web sites (or types of web sites), or even in other environments or settings separate from the Internet. If that type of data is available and accessible, then the invention could make effective use of that data in performing its predictive functions.

Also, it would be particularly advantageous to have included in the first data stream 20 information that identifies the personal attributes of the user being observed (e.g., age, gender, occupation, etc.), to the extent that such information has been previously inferred or compiled for that user or can be gathered during the on-line session being observed.

The AM 22 compares the contents of the two data streams 20, 26 to identify points of similarity between (a) the on-line activities of the current user being observed, i.e., the web sites (or types of sites) visited by that user and the time spent at each, and (b) the web sites (or types of web sites) visited by the other users and the average time spent at each. (Clearly, if the data stream 20 also contains information indicative of the personal attributes of the current user, this information will be factored into the identification of points of similarity.) Based on the existence of a discernible behavioral pattern exhibited by multiple "similar" users, i.e., users who had visited the same sites (or types of sites) for comparable amounts of time, the AM 22 is able to generate an output that attributes to the current user this same behavioral pattern.

The determination of a likely behavioral pattern can be facilitated by the addition of an optional third stream of data that identifies, and distinguishes between, all possible activities in which the current user can choose to participate. A particularly useful way of doing so within the context of the described Internet-based on-line transactions, is to provide to the AM 22 a third data stream 30 from a database 32 that identifies all the separately identifiable Internet web sites, categorized by common subject matter according to a predetermined set of criteria, that are available to the typical user of the given ISP. Then, the AM 22 selects those sites that the current user is most likely to visit during the present on-line session, based on his/her similarity to prior users that visited those same sites, and generates an output indicative of those selected sites.

The output of the AM 22 is delivered to a content delivery system (CDS) 34 that is configured to permit the web site to deliver advertisements or other informational content to visitors to that web site; an example of such a CDS is described in U.S. patent application Serial No. 09/558,755. Although the CDS 34

is shown as being resident on the master server 18 that supports the other components of the present invention, the invention would still function effectively if the CDS 34 were resident elsewhere, e.g., on the web server 12 that supports the operation of the particular web site being visited. In other words, the CDS could be one that is integrated with the other components of the invention in a "stand-alone" system and that is linked to, and accessible by, multiple web sites, or it could be one whose use is dedicated exclusively to a single web site. Based on well-known operations research and yield management techniques, the CDS 34 selects the most appropriate informational content (i.e., appropriate in light either of the pattern of behavior attributed to the current user or of the types of web sites deemed most likely to be visited) that is currently available to the CDS 34, for delivery to the current user and presentation at his/her client machine 10 via the CDS software client 24 installed in the client machine. Thus, for example, knowing that the current user has a greater than 50% chance of visiting certain categories of web sites within the next several minutes would permit the CDS 34 to choose a banner advertisement or other informational message from its current inventory that is most closely aligned with the interests served by such web sites. These messages then can be delivered to the present user in light of his/her anticipated behavior or preferences. Alternatively, the CDS 34 could direct the current user to the actual web sites that have been identified as most likely to be compatible with that user's preferences.

It should be pointed out that even if the database 28 showing the historical activities of other users were to be empty, or if for some reason the second data stream 26 should be unavailable, the present invention would still be able to operate. As the current user participates in activities in real time during his/her current on-line session, i.e., visits various web sites, this cumulative activity can be captured by the CDS software client 24. At the same time, CDS software

clients 24 on multiple other client machines 10 can capture the contemporaneous activities of other Internet web surfers, to identify the sites or types of sites that they are visiting in parallel with the current user's visits. Data reflecting all this activity can be used to populate database 28 and so create that an ad hoc historical archive. Until some critical mass is achieved, i.e., until there is a sufficient amount of data to support a credible determination of behavioral patterns for the current user, informational content would be delivered according to some generic default criteria previously established for the CDS 34. However, as time passes, the ad hoc historical archive would become sufficiently detailed to permit the generation of the second data stream 26.

The flow chart of FIG. 3 demonstrates the sequence of events in the operation of the preferred embodiment of the present invention. At step 40, a determination is made that a particular user has initiated an on-line session that is to be monitored. At step 42 the collection of data indicative of that user's on-line activities is initiated, and at step 44 the data stream reflecting (a) the cumulative activities and attendant conditions, and (b) the behavioral patterns of the other users is received. Step 46 compares the current user's activities with those of the other users. If, from a probability and statistical perspective, step 48 determines that there is not a sufficient number of similarities (based on predetermined criteria) to support a meaningful prediction of the current user's behavior, then no prediction will be made. However, at step 50 it is possible nevertheless to deliver targeted informational content to the current user based on predetermined criteria unrelated to the present situation, and at step 52 to add data reflecting the current user's activities to the database of historical activity so that over time, in combination with data reflecting the activities of other users, step 48 will yield sufficient similarities on which to base a meaningful prediction.

If, on the other hand, step 48 initially yields sufficient similarities to justify a prediction, then at step 54 a pattern of behavior is attributed to the current user that is consistent with the patterns of behavior previously input at step 44. Thereupon, at step 56 target informational content can be delivered to the current user (either directly or, e.g., via a web site that the user is visiting) in accordance with the attributed pattern of behavior. Finally, if step 58 determines that the current user is still involved in his/her on-line session, then the process can resume at step 42, but this time using a more recently updated historical activity input. Otherwise, the prediction activity with respect to the current user is terminated at step 60.

As illustrated in FIG. 4, in accordance with an alternate embodiment of the invention, user behavioral and preference information is developed and stored on the client machine 10' itself rather than on the master server 18 as in previously described embodiments. In accordance with the FIG. 4 embodiment, a CDS client 24' determines user behavioral or preference information by analyzing the online activities of the user including, for example, (a) the URLs requested by the user and (b) the amount of time that the user has spent investigating requested web sites. These activities are analyzed in connection with generally known data on profiles of visitors at particular web sites. The user behavioral or preference information is sent by the client machine 10' via the POP server 16 directly to a web site being visited using, for example, cookies. The web site can then dynamically customize content to be delivered to the user.

Although the present invention has been presented in the context of an Internet-based interactive transaction, it can operate just as effectively in other environments. For example, the client machines 10 could be replaced by television receivers configured with set-top boxes or comparable devices, whether fully interactive or not, that permit some degree of two-way

communications between a television viewer and a television system head end, and whether connected by land-based cables, by a satellite hook-up or by other means. Such devices typically include a mechanism for identifying a television viewer by some unique designator (e.g., by the set top box serial number) and for
5 collecting information corresponding to his/her channel selections. This compiled information on channel-viewing activity then can be processed in accordance with the teachings of the present invention. In the case of a fully interactive television system, other activities in which the viewer engages, in addition to merely watching television programs, can similarly be monitored and
10 recorded for processing by the present invention.

Similarly, the client machines 10 could be replaced with any of a variety of consumer information appliances, i.e., individually addressable devices connectable to a network (either wired or wireless) that have a user interface and
15 which are capable of storing and executing programs and of displaying information.

Also, although the disclosed embodiment has been shown as implemented in software operable on one or more computer systems, it could be
20 effectively implemented in hardware components.

While specific embodiments of the invention have been described herein, it will be appreciated by those skilled in the art that other modifications and variations, including additions and subtractions, may be made to the disclosed
25 invention without departing from its spirit and scope, and it is intended that such modifications remain within the scope of the following claims.